2024 IEEE International Conference on
Autonomous Robot Systems and Competitions (ICARSC)
Paredes de Coura, Portugal – May 2-3, 2024

2024 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC) | 979-8-3503-5234-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICARSC61747.2024.10535936

Leveraging Multi-Object Tracking in Vision-based Target Following for Unmanned Aerial Vehicles

1st Diogo Ferreira
Institute for Systems and Robotics
Instituto Superior Técnico, UL
Lisbon, Portugal
diogocostaf@tecnico.ulisboa.pt

2nd Meysam Basiri
Institute for Systems and Robotics
Instituto Superior Técnico, UL
Lisbon, Portugal
meysam.basiri@tecnico.ulisboa.pt

***Abstract*—This paper presents an autonomous Vision-based Mobile Target Tracking and Following system designed for Unmanned Aerial Vehicles, enabling them to effectively track and pursue mobile targets on the ground. Rather than the conventional Single-Object Tracking strategy, this paper explores the capabilities of recent advances of Multi-Object Tracking methods in a target following scenario. We leverage cutting-edge detection and tracking capabilities by integrating You Only Look Once (YOLOv8) with the MOT BoT-SORT, enabling the extraction of multi-target information. Complementing visual data, we incorporate a depth sensing module to enhance distance estimation when feasible. For target following, a 3D flight control system is proposed, capable of reacting to the targets speed and direction changes, while maintaining line-of-sight. The full system is deployed using the MRS UAV System [1] in a realistic simulation capable of future real-world deployment. Results show that the system achieves precise target tracking and following, resilient to partial and full occlusions in dynamic environments, effectively distinguishing the followed target from bystanders.**

***Index Terms*—UAV; Multi-Object Tracking; YOLOv8; BoT-SORT; Flight Control; MTT**

I. Introduction

Vision-based Mobile Target Tracking (MTT) for Unmanned Aerial Vehicles (UAV)s is an ever-evolving field of operations, essential in several areas that benefit from an airborne source of information such as: crowd management [2], ground [3] or oceanic [4] vehicle management, search and rescue missions [5] and auto-landing applications [6].

The problem of Vision-based MTT and Following usually encompasses the detection of the object, followed by a target tracking algorithm that monitors the position of the object over time in the image. Finally, the relative position of the object in relation to the UAV is computed to give feedback to the controller and allow target following. Target tracking strategies fall into two primary categories: Single-Object Tracking (SOT) approaches which focus on precisely detecting, tracking and following a single predetermined target in the image [7]; and Multi-Object Tracking (MOT) approaches that keep track of multiple targets, handling more complex scenarios [3], [4].

This work was partially supported by LARSyS funding (DOI: 10.54499/LA/P/0083/2020, 10.54499/UIDP/50009/2020 and 10.54499/UIDB/50009/2020) and Aero.Next project (PRR - C645727867-00000066)

Many works [8]–[12] have been developed in the field of Vision-based MTT and Following by UAVs using SOT methods, namely the Kernelized Correlation Filter (KCF) algorithm. Correlation-based trackers such as KCF propose one-shot learning and show good performances without GPU acceleration, which makes them very appealing for embedded systems with computational limitations [13]. However, KCF relies on the appearance model of the object being tracked. When occlusions occur, significant changes in object appearance can hinder accurate tracking, potentially leading to tracking failures. Hence, the KCF method is very susceptible to partial target occlusions and can only track the target in the image after it is selected in the first frame, which is usually done manually [8], [9], [12]. To partially tackle the problem of recovery after occlusion, an algorithm was developed in [9] that analyses the motion between frames to detect movement indicative of the target. However, this can be susceptible to noise and dynamic environments. Another common approach is to use the Kalman Filter (KF) in conjunction with the KCF [11], [12]. Reference [11] also takes advantage of deep learning methods (YOLOv3) to initialize the KCF tracker and redetect the target after a full occlusion which can be done if there are no similar targets in the image.

On the other hand, due to their complexity, MOT methods have more difficulties handling camera motion and view changes provoked by the UAV movement but work in dynamic multi-target environments. Recently, some works attempt to address this issue by improving camera motion models [14], [15]. MOT applications have the advantage of working under a tracking-by-detection approach which performs a detection step followed by a tracking step, in every frame, instead of the single detection step at the beginning. The MOT approach allows for the consistent use of new deep learning-based detection methods (such as YOLOv8 [16]) in order to increase the reliability of the system as a whole. Reference [17] demonstrates the accuracy of a tracking-by-detection approach using the previous YOLOv7 version and the state-of-the-art BoT-SORT algorithm. Results show an effective solution for target occlusion and identity switching in pedestrian target tracking, even under poor illumination conditions and complex scenes. Reference [18] applies a similar system that utilizes YOLOv8 and BoT-SORT within a Synthetic Aperture Radar

(SAR) imaging framework. The experimental findings suggest that the proposed method exhibits high precision in both detection and tracking, with real-time capabilities.

To our knowledge, despite recent advancements in MOT, there have been no practical applications of these techniques in Vision-Based MTT and Following from a UAV. Our research aims to fill this gap by exploring the potential of MOT in target following, addressing limitations in current SOT approaches. This approach opens avenues for new applications such as conditional target switching, crowd following, and enhanced redetection methods.

In mobile target following, the deviation between the target's pixel position and the image frame's center can guide the camera towards the target, achieving line-of-sight (LOS) [8]–[11]. While some studies use this deviation to steer a gimbal towards the target [8], [9], others utilize it for lateral and vertical movements to maintain focus on the target [10], [11]. Once in LOS, maintaining the relative distance for target following becomes crucial. Some methods estimate distance using the standard pinhole imaging model combined with an Extended Kalman Filter (EKF) to handle observation noise [9], while others use the proportion of the target's size in the image as a reference for following speed [8], [11].

Control algorithms for mobile target following vary across studies, from direct 3D position inputs [10], [11] to the use of proportional navigation (PN) with cascade PID controllers [8]. Other works use a switchable tracking strategy based on estimated distance, transitioning between observing and following modes with different control methods [9]. Many existing distance estimation methods suffer from limitations in precise target positioning, highlighting the need for more accurate methods. Incorporating a depth sensing module, when feasible, presents a promising solution to address this issue, as suggested by [8].

The system proposed in this paper aims to make the following contributions to the research field of Vision-based MTT and Following using UAVs: 1) Implement a Visual Detection and Tracking system based on state-of-the-art YOLOv8 detection algorithm and the multi-object tracker BoT-SORT algorithm [15] in a scenario of MTT and Following. 2) Design a 3D flight control algorithm capable of following a target in a dynamic environment using RGB information coupled with depth information to achieve better results. 3) Create an innovative redetection algorithm utilizing the capabilities of multi-target information to handle partial and full target occlusions in dynamic environments. 4) Test the full system in a realistic simulation environment utilizing the capabilities of ROS, Gazebo and the MRS-UAV System [1].

The rest of the paper is organized as follows: In Section II, the system configuration is introduced. In Section III, the Visual Detection and Tracking Module is exposed. Section IV explains the distance estimation methods. Section V explains the high-level Following Flight Controller and Section VI goes over all the states and transitions in the System Mode Switcher. Experimental results are presented in Section VII. Concluding remarks and future work are discussed in Section VIII.

II. System Configuration

The system configuration is shown in Fig. 1.

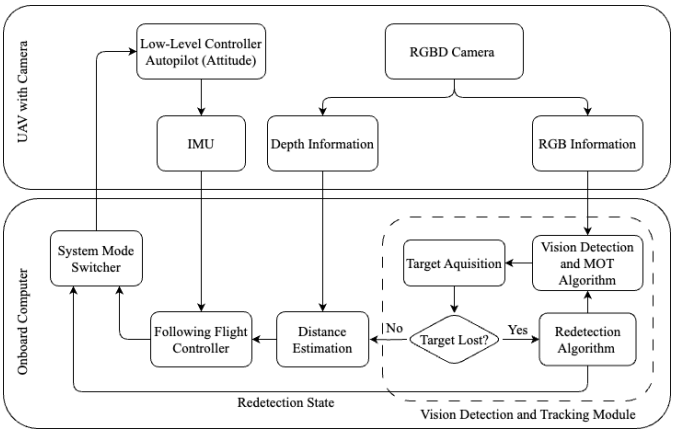


Fig. 1. System Configuration of the vision-based detection, tracking and following system

The "Vision Detection and MOT Module" is responsible for using the RGB data from the camera and performing the detection and tracking of all mobile targets in the image, including the target determined to be followed. It is also responsible for clearly differentiating between the target and the bystanders during normal operations or in the event of occlusions.

The output of this module is sent to the "Distance Estimation" stage to get an estimate of the relative distance from the UAV to the target as control error for the "Following Flight Controller". The distance estimation may also be obtained from a depth sensor, such as ones included in a RGBD camera. The controller then computes the velocity and heading commands for the UAV and sends them to the "System Mode Switcher". In turn, it will determine which inputs are sent to the autopilot according to the current state of the system. The autopilot controls the attitude of the UAV and sends the IMU data as feedback to the controller. The autopilot of choice is the MRS UAV System [1] SE(3) controller, which feeds directly into a Pixhawk board. Path planning methods could also be integrated if required to generate efficient trajectories while avoiding obstacles [19].

In this study, the algorithm will undergo testing in complex scenarios featuring multiple pedestrians in motion, where challenges like occlusions and disturbances are likely to arise.

III. Visual Detection and Tracking Module

A. Object Detection and Tracking

The first step to achieve target following is to detect the target in the image and assign it a specific ID to track it over time. This is done by using a track-by-detection approach where we use a pre-trained object detector to identify target objects within individual frames. These detections are then linked over time by the tracker, by attributing unique identifiers for each object. To achieve this setup, we take advantage

of the state-of-the-art detector YOLOv8 and combine it with the BoT-SORT MOT algorithm [18]. YOLOv8 provides the detections via bounding boxes to the tracker which in turn performs data association to match each detection with a corresponding ID.

B. Target Acquisition

After take-off and initialization of the detection and tracking algorithm, users have the option to select the target to follow from the list of detected people. In this work, the algorithm designates the first person detected as the target to follow, storing its respective ID for subsequent frames. In order to take full advantage of the MOT capabilities, the position of other people in the image and respective assigned IDs will also be stored to enhance the redetection capabilities in case of target occlusion.

After detection, a KF inspired by [12] is used to predict the position of the target in the image even after occlusion. The states used are the center coordinates of the bounding box (c_x, c_y), the size of the bounding box (width - w, height - h) and the speed of movement in the image (v_x, v_y), calculated from the movement of the center. The states are defined in (1).

$$s(t) = (c_x,\, c_y,\, v_x,\, v_y,\, w,\, h) \tag{1}$$

The observations are shown in (2).

$$o(t) = (c_x,\, c_y,\, w,\, h) \tag{2}$$

The KF updates with new observations every time the target is visible. When the state is set as "Target Missing", the predictions from the KF are used to continue following since it is likely to reappear in the next few frames.

C. Redetection algorithm

In vision-based detection and tracking of mobile targets from UAVs, occlusions are frequent, especially in crowded areas. Redetection algorithms are crucial for reliable operation, particularly in scenarios involving multiple individuals. To begin the target following process, the algorithm verifies detections using assigned IDs from the YOLOv8+BoT-SORT setup. If the assigned target ID is absent in the detections, three possible situations may arise:

1) **Target ID change:** the target is still visible in the image and identified, however, the MOT algorithm assigned it a different ID.
2) **Target missing:** if no ID change is detected, then the target is assumed missing and a counter is started. During this step, we continue to follow an estimate of the position of the target given by the KF.
3) **Target lost:** after some consecutive missing frames, the counter reaches zero and the target is assumed lost/hidden. The UAV will hover and search for the target until it is redetected.

To perform the redetection process, the algorithm searches all detections for possible candidates. Firstly, it will attempt to check if a "Target ID Change" is in place. If no match is found, then it will enter the "Target Missing" state and later the "Target Lost" state.

A viable candidate for a successful redetection must fulfill the following conditions: 1) It must represent a new detection not previously tracked, thereby excluding individuals already accounted for as they cannot be the target. 2) Candidate detections are evaluated based on a minimum interception threshold with the latest known position of the missing target, determined using an Interception Over Union (IOU) approach. Bounding boxes are scaled to enhance recovery. 3) Among all the candidates that fulfill step 1 and step 2, it must be the one that scored the highest on step 2.

IV. Distance Estimation

Distance estimation is a crucial step to be able to effectively follow the target. In this case, this value can be roughly estimated based on the detection bounding box provided by the "Visual Detection and Tracking Module" or more accurately obtained by a depth sensor installed. Despite the preference for depth information, it is not always available in most configurations.

In order to correctly estimate distance (d in Fig. 2), we use a relation between the pixel height (h) of the target in the image and a tuned constant value C (3).

$$d = \frac{C}{h} \tag{3}$$

The height of the target in the image is chosen as reference over the area occupied by the bounding box ($w * h$) [12] because it mainly depends on the real height of the target and the distance to the target. On the other hand, the width of the target in the image may vary according to the direction of movement and position of the arms and legs. The constant value can be previously tuned for the average height of a human and adjusted during operations after getting a better estimate from on-board depth sensors. In non-person target tracking and following (such as vehicles or robots), the area of the bounding box can be considered.

The distance estimation can be susceptible to observation noise. To achieve smooth control, an exponential low-pass filter in a discrete-time system is used as shown in (4). d_{low_pass}

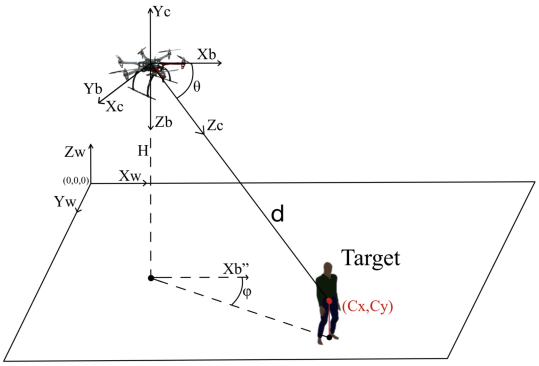


Fig. 2. Relative distance to the target d and reference frames - UAV body frame F_b, Camera frame F_c and World frame F_w. θ represents the camera pitch angle. H represents the altitude of the UAV. X_b'' is the parallel line to X_b in the plane. ϕ represents the heading of the UAV.

is the filtered value, $d(t_{k-1})$ is the previous estimation and $d(t_k)$ is the current estimation. The α value is tuned to prevent higher frequency oscillations in the measurements that could disrupt the controller.

$$d_{low\_pass} = (1-\alpha) * d(t_{k-1}) + \alpha * d(t_k),\ \alpha \in [0,1] \quad (4)$$

V. Following Flight Controller

The Following Flight Controller is composed of three separate controllers to achieve accurate 3D following of the target: one for the heading, one for the altitude (vertical velocity) and a third one to control the horizontal velocity of the UAV.

This work draws inspiration from [10], [11], employing the deviation between the pixel location of the target and the central position of the image frame as feedback for maintaining the UAV's focus on the target. The heading and altitude controllers are proportional controllers designed to center the target in the image. Normalized references for the horizontal, and vertical pixel positions are used (5). The principle point coordinates ppx and ppy are obtained from the intrinsic parameters of the camera.

$$\begin{aligned} X_N &= \frac{c_x}{ppx} - 1 \\ Y_N &= \frac{c_y}{ppy} - 1 \end{aligned} \quad (5)$$

The heading reference for the controller and vertical velocity are then computed using (6) and (7) respectively.

$$\phi_{ref} = \phi_{current} + K_\phi * X_N \quad (6)$$

$$V_z = K_H * Y_N \quad (7)$$

The algorithm uses an aim and approach strategy similar to the PN strategy in [8] to achieve target following, by using the heading and altitude controllers to aim and the horizontal velocity controller to approach the target. The control output for the velocity is calculated by using a PI controller taking the estimated distance as input. The horizontal velocity controller error is defined according to (8).

$$error = d - d_{desired} \quad (8)$$

Where the desired distance $d_{desired}$ is defined by the user. In order to achieve smooth control, the velocity value V goes through a slew rate limiter that prevents sudden and aggressive maneuvers which may cause loss of line of sight. The slew rate limiter also sets the initial control output as zero, allowing for controlled initial movement. The limited rate of change SR is defined according to (9).

$$SR = \frac{V(t_k) - V(t_{k-1})}{t_k - t_{k-1}} \quad (9)$$

The velocity controller output is sent to the System Mode Switcher in the world frame, by multiplying the limited velocity value V_{SR} by the heading show in (10).

$$\begin{aligned} V_x &= cos(\phi) * V_{SR} \\ V_y &= sin(\phi) * V_{SR} \end{aligned} \quad (10)$$

VI. System Mode Switcher

The System Mode Switcher is the ultimate responsible for providing the low-level controller with the appropriate control references based on the state of the system.

The system can be in the following states:

- **Search Mode:** no target has been detected and identified yet.
- **Adjusting Mode:** only the heading and altitude controllers are used to center the target in the image;
- **Following Mode:** this may be defined as the mode for standard operations, when the target is identified and being followed in 3D. This is also the chosen mode if a Target ID Change occurred.
- **Target Missing Mode:** as previously defined in Section III-C, the target is not visible in the image;
- **Target Lost Mode:** also defined in Section III-C, if the target has been missing for several consecutive frames it is assumed lost/hidden.

Fig. 3 presents a flowchart representing the system modes and their respective interactions.

Following take-off and systems check, the UAV will enter "Search Mode". In this mode, it will perform a climb until the maximum defined safe altitude and slowly rotate until a target is detected.

After the target is detected and identified, the UAV will enter "Adjusting Mode", adjusting the target to the center of the image as best as possible, considering altitude safety limits. This allows the system to have a more controlled first approach to the target. After the "Adjusting Mode" timer is over, the system enters regular operations with the "Following Mode", which feeds the output from the controllers to the autopilot.

Regarding the first recovery mode - "Target Missing Mode", the System Mode Switcher will continue to send the commands from all three controllers however, the estimated distance and subsequently the error will be calculated using the predictions from the KF.

Finally, in the second recovery mode - "Target Lost Mode", the System Mode Switcher will set the UAV to hover. This is done to prevent further deviation from the lost target, here assumed to be hidden behind some obstacle. The system will remain in this mode until a successful redetection is obtained. Unlike the "Target Missing Mode" that directly shifts to "Following Mode", here, once the target is redetected, the System Mode Switcher will initiate the "Adjusting Mode" for a brief period and reset the rate limiter and the low pass filter previous values. This will ensure the target is properly redetected and followed.

VII. Experiments

The system is implemented and tested using the MRS UAV System [1], which provides valuable data for future real-world deployment. The experimental setup is a Hexacopter DJI-F550 equipped with RGBD camera Realsense F435, while low-level control and IMU acquisition are managed by a Pixhawk running PX4 firmware. Simulations are conducted within Gazebo-ROS.

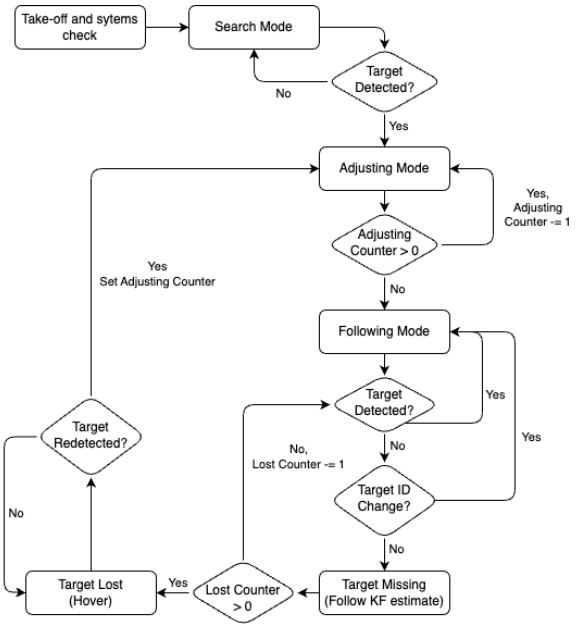


Fig. 3. Flowchart of the System Mode Switcher

TABLE I
DATA FROM THE 5 RANDOMIZED EXPERIMENTS

Test Nº	UAV Distance	Target Distance	Visual Accuracy	Estimation Accuracy	Depth Use
1	$92.5m$	$118.6m$	99.51%	$0.504m$	67.23%
2	$71.3m$	$111.2m$	99.72%	$0.455m$	69.46%
3	$77.5m$	$104.7m$	97.10%	$0.646m$	61.35%
4	$85.4m$	$114.6m$	98.53%	$0.570m$	64.85%
5	$80.1m$	$93.74m$	99.56%	$0.584m$	71.51%

To assess system performance in open environments with multiple pedestrians, five tests are conducted in a 900-square-meter area, featuring 3 to 6 pedestrians walking at an average speed of 1 to 1.5 meters per second. Each test spans approximately 2 minutes, with randomized and not pre-set trajectories.

The desired distance ($d_{desired}$) is set to 11 meters, considering the Realsense's depth range (10 meters) with an added 1-meter margin to enhance responsiveness to sudden target movements toward the UAV. This configuration proves effective, with depth values utilized for over 60% of the test duration across all five trials. Combined distance estimation methods leveraging target size and depth values demonstrate an accuracy of approximately 0.5 meters during active target tracking.

Results are summarized in Table I, where average distance estimation error is measured relative to ROS-Gazebo provided distances. "Visual Accuracy" denotes the percentage of frames where the target is correctly identified, while "Depth Use" represents the percentage of time the depth sensor data is utilized for position estimation. UAV and Target Distances refer to the total travel distance for each test. Across the five tests, the UAV successfully tracked and followed the target, covering a total travel distance of 542.84 meters.

To evaluate the capabilities of the system to maintain LOS and keep the target in the center of the image, Fig. 4 illustrates the density distribution of the target's bounding box center position in the image. The target predominantly remains centered, although deviations are more pronounced along the vertical axis due to the lack of gimbal stabilization for the camera, coupling forward/backwards movement with a downward/upwards pitch maneuver.

To further evaluate the full system, an experiment was conducted that includes partial and full occlusions up to 10 seconds. Fig. 5 depicts the sequence of 4 frames: preceding the target's concealment, during the "Target Lost" phase, upon target reappearance, and subsequent to target redetection. The velocity output is shown in Fig. 6. The estimated distance vs real distance is shown in Fig. 7. Other results from the experiment are shown in Table II.

Overall, results align closely with those of obstacle-free experiments, with the only notable difference being a decrease in "Visual Accuracy." This decrease can be attributed to instances where the target experiences full occlusions during the course of the experiment. Velocity values are kept within reasonable ranges throughout the experiment. The distance to the target is correctly estimated with an average error of 0.599 meters with some higher error when the target changes direction. The distance to the target is kept around the desired distance of 11 meters with a 60.72% depth use. A demonstration video showcasing the entire 10-minute experiment using our system is available at the following link: https://youtu.be/n7PUNkULv9g.

VIII. CONCLUSIONS

In this work we developed a vision-based mobile target tracking and following algorithm for unmanned aerial vehicles. The system proposed uses an innovative strategy of deploying a Multi-Object Tracking algorithm in a target following scenario enhancing redetection capabilities. Results show that the

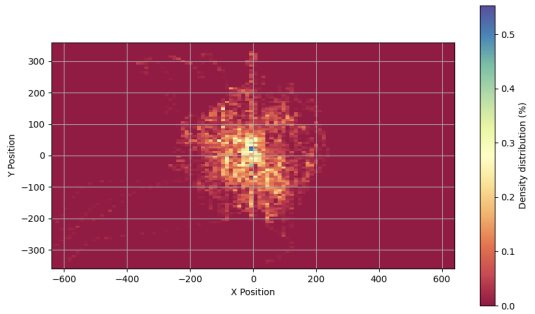
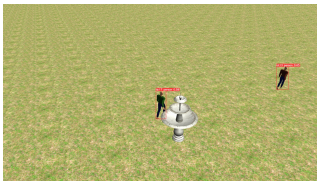


Fig. 4. Density probability distribution of the center of the target in the image frame during the 5 tests.

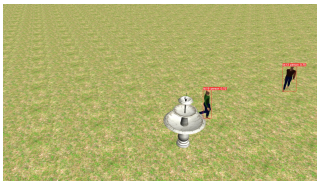
TABLE II
DATA FROM THE EXPERIMENT INCLUDING TARGET FULL OCCLUSION

Time (m:s)	UAV Distance	Target Distance	Visual Accuracy	Estimation Accuracy	Depth Use
$9m35s$	$366.8m$	$504.2m$	93.28%	$0.599m$	60.72%

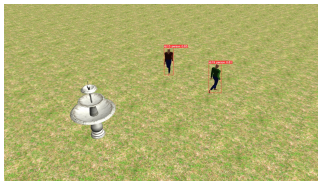


(a) Following the target (left). Bystander in the distance (right).

(b) Target (left) hidden behind obstacle.



(c) Target reapears in the detections.



(d) Target (right) redetected and followed.

Fig. 5. Redetection algorithm when the target is hidden behind an obstacle.

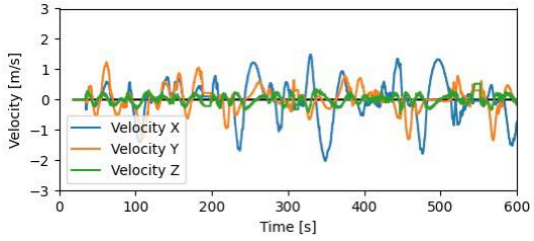
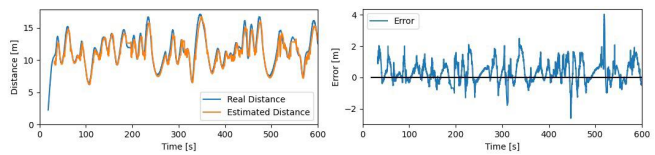


Fig. 6. Velocity in the world frame X (blue), Y (orange) and Z (green).

system is capable of handling complex and dynamic where partial or full occlusions are common. Moreover, the 3D flight controller effectively follows the target while keeping it centered in the image showcasing robustness even during sudden direction changes. The complete system was tested with the MRS UAV System [1] which makes for a solid foundation for future real-world implementation. Hence, our work not only contributes to the advancement of Vision-Based Mobile Target Tracking and Following but also sets the stage for future exploration and application of MOT techniques in UAV operations. Use cases of the developed approach include not only the improvement of redetection methods as show in this work, but also the possibility of changing the followed target during execution without reinitializing the algorithm.

Current limitations include the maximum velocity and direction changes of the target, which may cause loss of LOS if there is a sudden and fast movement towards the UAV, since the tracking distance is set to be as close to the person as possible for enhanced depth accuracy. In cases where these are expected, a larger follow distance might need to be considered. Future research will focus on exploring real-world scenarios using the developed system.



(a) (b)

Fig. 7. Comparison between the real and estimated distance from the target.

REFERENCES

[1] T. Baca, M. Petrlik, M. Vrba, V. Spurny, R. Penicka, D. Hert, and M. Saska, "The mrs uav system: Pushing the frontiers of reproducible research, real-world deployment, and education with autonomous unmanned aerial vehicles," *Journal of Intelligent & Robotic Systems*, vol. 102, p. 26, 2021.

[2] H. Hou, C. Shen, X. Zhang, and W. Gao, "Csmot: Make one-shot multi-object tracking in crowded scenes great again," *Sensors*, vol. 23, no. 7, 2023.

[3] X. Liu and Z. Zhang, "A vision-based target detection, tracking, and positioning algorithm for unmanned aerial vehicle," *Wireless Communications and Mobile Computing*, vol. 2021, p. 5565589, 2021.

[4] Y. Jie, L. Leonidas, F. Mumtaz, and M. Ali, "Ship detection and tracking in inland waterways using improved yolov3 and deep sort," *Symmetry*, vol. 13, no. 2, 2021.

[5] J. Qi, D. Song, H. Shang, N. Wang, C. Hua, C. Wu, X. Qi, and J. Han, "Search and rescue rotary-wing uav and its application to the lushan ms 7.0 earthquake," *Journal of Field Robotics*, vol. 33, no. 3, pp. 290–321, 2016.

[6] J. Morales, I. Castelo, R. Serra, P. U. Lima, and M. Basiri, "Vision-based autonomous following of a moving platform and landing for an unmanned aerial vehicle," *Sensors*, vol. 23, no. 2, 2023.

[7] Z. Soleimanitaleb and M. A. Keyvanrad, "Single object tracking: A survey of methods, datasets, and evaluation metrics," 2022.

[8] X. Liu, Y. Yang, C. Ma, J. Li, and S. Zhang, "Real-time visual tracking of moving targets using a low-cost unmanned aerial vehicle with a 3-axis stabilized gimbal system," *Applied Sciences*, vol. 10, p. 5064, 07 2020.

[9] H. Cheng, L. Lin, Z. Zheng, Y. Guan, and Z. Liu, "An autonomous vision-based target tracking system for rotorcraft unmanned aerial vehicles," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 1732–1738.

[10] Y. Feng, D. Wang, and K. Yang, "Research on target tracking algorithm of micro-uav based on monocular vision," *Journal of Robotics*, vol. 2023, p. 6657120, 2023.

[11] D. Luo, P. Shao, H. Xu, and L. Wang, "Autonomous following algorithm for uav based on multi-scale kcf and kf," in *2023 4th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, 2023, pp. 430–436.

[12] H. Wei, "A uav target prediction and tracking method based on kcf and kalman filter hybrid algorithm," in *International Conference on Consumer Electronics and Computer Engineering*, 2022, pp. 711–718.

[13] S. Yadav and S. Payandeh, "Critical overview of visual tracking with kernel correlation filter," 12 2021.

[14] S. Liu, X. Li, H. Lu, and Y. He, "Multi-object tracking meets moving uav," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 8866–8875.

[15] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," 2022.

[16] M. Hussain, "Yolo-v1 to yolo-v8, the rise of yolo and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol. 11, no. 7, 2023.

[17] T. Li, Z. Li, Y. Mu, and J. Su, "Pedestrian multi-object tracking based on YOLOv7 and BoT-SORT," in *Third International Conference on Computer Vision and Pattern Analysis (ICCPA 2023)*, L. Shen and G. Zhong, Eds., vol. 12754, International Society for Optics and Photonics. SPIE, 2023, p. 127541I.

[18] S. Yan, Y. Fu, W. Zhang, W. Yang, R. Yu, and F. Zhang, "Multi-target instance segmentation and tracking using yolov8 and bot-sort for video sar," in *2023 5th International Conference on Electronic Engineering and Informatics (EEI)*, 2023, pp. 506–510.

[19] M. A. Basiri, S. Chehelgami, E. Ashtari, M. T. Masouleh, and A. Kalhor, "Synergy of deep learning and artificial potential field methods for robot path planning in the presence of static and dynamic obstacles," in *International Conference on Electrical Engineering (ICEE)*, 2022, pp. 456–462.